IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re PATENT application of	)
Yeogirl YUN et al.	) Confirmation No. 3718
Serial No. 09/287,296	) Group Art Unit: 3624
Filed: April 7, 1999	) Examiner: T. T. Havan
For: METHOD AND APPARATUS FOR	)
DEFINING DATA OF INTEREST	)

APPEAL BRIEF

Mail Stop Appeal Brief- Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Appeal Brief is submitted in support of the Notice of Appeal filed April 5, 2005.

I.   REAL PARTY IN INTEREST

CNET Networks, Inc. is the real party in interest.

II.   RELATED APPEALS AND INTERFERENCES

There are presently no appeals or interferences known to the Appellants, the Appellants' representative, or the assignee, which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III.   STATUS OF THE CLAIMS

Claims 1, 2, 4-7, 9-21 and 23-34 are pending, as submitted in an amendment filed on December 2, 2004 in response to the final Office Action mailed October 5, 2004. Claims 3, 8, and 22 were previously canceled. The present case has been more than twice rejected. This Appeal is taken from the rejection of claims 1, 2, 4-7, 9-21 and 23-34, the claims being submitted in the APPENDIX herewith.

IV.   **STATUS OF AMENDMENTS**

Amendment After Final was submitted on December 2, 2004. In the Advisory Action mailed May 10, 2005, the Examiner maintained his rejection of the application, but agreed to enter the amendments to the claims for the purposes of appeal. The Examiner further provided an explanation of how the amended claims would be rejected in the Advisory Action.   Thus, claims as submitted in the Amendment After Final, which are set forth in the APPENDIX, are the subject of appeal.

V.   **SUMMARY OF THE INVENTION**

The prior art fails to provide a method and system that facilitates extraction of data of interest from a plurality of websites. In the prior art, crawlers are created by computer programmers to retrieve information from a particular web site, for example, to extract desired information for a category of products from on-line merchants for use in an electronic catalog. However, different web sites, for example, web sites of different on-line merchants, utilize different data structures. There is no standardized structure, method or protocol for presenting and storing information or data among different web sites that is uniformly followed by different on-line merchants. In addition, each web site generally utilizes a plurality of web pages in the web site to which a user has to navigate to obtain the desired data of interest regarding a product available through the web site, for example.

A crawler that is created and used to extract data from one web site generally cannot be used to extract data from other web sites due to the variations in data structure, method and/or protocol implemented by other web sites. Thus, a new crawler must be created by a computer programmer to extract data for each web site, the creation of new crawlers being time consuming and expensive. Consequently, extracting data of interest, for example, regarding a particular product from a plurality of different web sites such as merchant web sites, can be extremely difficult, expensive, and time consuming.

The present invention provides a novel method and system for extracting data of interest from a plurality of web sites that greatly facilitates the extraction process

by providing tools that can be used, even by non-programmers, to extract desired information from the plurality of web sites. (See Pg. 3, lines 9-10; Pg. 8, lines 19-23). More specifically, the present invention allows the user to generate extraction patterns directly from the output of the web site itself, such as the HTML source view of a web browser, so that other desired information can also be extracted from the web site using the generated extraction patterns. (See Pg. 3, lines 13-20; Figs. 10, 14-18 and related disclosure in Pg. 23, lines 13-19; Pg. 24, line 18-Pg. 26, line 6).

Accordingly, one aspect of the present invention is directed to a method of extracting data of interest from a plurality of web sites, the method comprising for each respective web site, creating a respective description of data of interest that identifies the web site; developing an extraction pattern based on output from the respective web site using a graphical user interface tool, the extraction pattern extracting information from the respective web site; and associating the developed extraction pattern with the respective description of data of interest for the web site. The method also includes receiving a value that can be used as an extraction parameter for the developed extraction patterns; and obtaining the data of interest by querying web sites in the plurality of web sites using the value and the extraction patterns associated with the respective descriptions of data of interest, wherein when the data of interest includes data of interest from at least two web sites of the plurality of web sites, the data of interest from the at least two web sites is provided. (See independent claims 1 and 18).

In addition, another aspect of the present invention is directed to a computer data signal comprising a software module for creating a description of data of interest, the software module including a set of operations for interactively developing an extraction pattern based on output of a target web site using a graphical user interface tool, the developed extraction pattern for obtaining data of interest from the target web site; a set of operations for receiving a selection of an instruction from a predefined set of instructions for inclusion in the description of data of interest; a set of operations for associating the extraction pattern with the instruction; and a set of operations for testing the instruction using the extraction pattern and the contents of a buffer, wherein the buffer includes a portion of the output of the web site associated

with the description of data of interest. The computer data signal further includes a software module for using the description of data of interest to obtain data of interest from the target web site when a value that can be used as an extraction parameter for the developed extraction pattern is provided. (See independent claim 21).

Furthermore, another aspect of the present invention is directed to a computer implemented method of obtaining data of interest from a plurality of web sites comprising developing a description of data of interest for each web site in the plurality of web sites based on output from the plurality of web sites using a graphical user interface tool that includes a web browser, each respective description of data of interest specifying an address for a corresponding web site in the plurality of web sites and each respective description of data of interest including an extraction pattern identifying at least a portion of the output of a web site and extracting user specified information from the corresponding web site; receiving a value that can be used as an extraction parameter for the developed extraction patterns; and obtaining the data of interest by querying web sites in the plurality of web sites using the value and the extraction patterns in the respective descriptions of data of interest. (See independent claim 32).

Thus, as described in the Specification of the application, the present invention can be used by an individual such as a programmer, or even a non-programmer, to generate extraction patterns easily based on the output, such as the HTML source code, of the web site itself. A value can then be used in conjunction with the developed extraction pattern to extract different data of interest from the particular web site. Correspondingly, the present invention allows facilitated extraction of desired data of interest from a plurality of web sites in a rapid, cost effective manner, without requiring a programmer to create a crawler for each web site from which data of interest is desired. Example implementation of the present invention and development of extraction patterns are most clearly shown in Figures 14 to 17, and the corresponding portions of the specification describing these figures in Page 24, line 18 to Page 25, line 24.

VI.   THE APPLIED REFERENCE

The applied reference is U.S. Patent No. 4,992,940 to Dworkin.

VII.   ISSUES

The issue on appeal is whether claims 1, 2, 4-7, 9-21 and 23-34 are unpatentable under 35 U.S.C. 103(a) over U.S. Patent No. 4,992,940 to Dworkin.

VIII.   GROUPING OF THE CLAIMS

Each claim of this patent application is separately patentable, and upon issuance of a patent would be entitled to a separate presumption of validity under 35 U.S.C. §282.   For convenience in handling of this Appeal, the claims will be addressed in groups, as follows:

Group I.   Claims 1, 2, 18, 20, 25, 26, 29, 30, 32, and 34.

Group II.   Claim 4.

Group III.   Claims 5, 6, 9, 10, 16, and 19.

Group IV.   Claims 7.

Group V.   Claim 11.

Group VI.   Claims 12-15, 17, 21, and 31.

Group VII. Claim 23, 24, 27, 28, and 33.

Thus, pursuant to 37 C.F.R. §1.192(c)(7), in this Appeal, the rejected claims will stand or fall together only within each group.

IX.   ARGUMENTS - GROUPS I - VII

In the Advisory Action mailed May 10, 2005, the Examiner maintained his rejection of claims 1, 2, 4-7, 9-21, and 23-34 under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 4,992,940 to Dworkin.   The Applicants respectfully disagree and contend that the Dworkin reference fails to disclose, teach, or otherwise suggest the present invention for the reasons set forth herein below.

Dworkin reference relates to a system and method for automated selection of equipment for purchase where the user selects a category of product or service, and

W656626.1

the user is provided with a template which gives various criteria for the product or service selected so that the user can fill in the template with specifications of the product or service desired. (See Abstract; Col. 3, lines 48-54; Col. 5, lines 43-50). Dworkin discloses that upon receiving input from the user as to one or more criteria of the template, the system searches a database for all products that fulfills the requirements/specifications inputted by the user. (See Abstract; Col. 2, lines 6-18; Col. 6, lines 11-15). In this regard, Dworkin notes that the database may be the equivalent of thousands of catalogs of individual suppliers. (See Col. 3, lines 66-68). Dworkin also discloses that the database includes information regarding products from a plurality of vendors or distributors within the selected category. (See Col. 1, lines 65-68; Col. 3, lines 63-68). The results of the search are displayed for the user identifying the products together with the vendor. (See Col. 2, lines 19-25).

Thus, the cited Dworkin reference is essentially a multi-vendor catalog or search engine that receives specification requirements for a product from a user, searches an extensive database for products that satisfy the specification requirements, and provides a search result which identifies the products in the database that satisfy the user specified requirements, in conjunction with the identity of the vendor.

A.   GROUP I

The rejection of independent claims 1, 18, and 32, and dependent claims 2, 20, 25, 26, 29, 30, and 34 of Group I based upon Dworkin is improper in that Dworkin fails to disclose, teach, or otherwise suggest the present invention as recited these claims.

It is initially noted that the invention described in Dworkin is substantially different from the present invention. In this regard, it is important to recognize that Dworkin assumes that the database disclosed is populated with information regarding products that are available from plurality of vendors. Dworkin is completely silent as to how this database disclosed in Dworkin is populated with product information and/or vendor information. Thus, Dworkin presumes that the databases of different vendors have been searched, desired product information extracted, and stored in the database so as to allow users to search the database based on user inputted criteria.

W656626.1

However, as explained, different web sites (such as web sites of different product vendors) utilize different data structures. In addition, each web site typically utilizes a plurality of web pages in the web site which would require the user to navigate through various web pages in order to obtain/extract the desired product data. Dworkin is silent as to how the disclosed database of information regarding products from a plurality of vendors can be provided, and populated with product data from the catalogs of numerous suppliers. Correspondingly, without explicit teaching to the contrary, one of ordinary skill in the art would understand that the conventional method of using crawlers would be implemented in the system and method disclosed in Dworkin. Consequently, Dworkin does not contribute to solving the problem of requiring individual, web site specific, crawlers to be created by a programmer to extract the product information required to populate the database used by the system and method of Dworkin. Again, the creation of new crawlers for each web site would be time consuming and expensive.

In contrast, the present invention provides a method and system for extracting the desired data of interest from a plurality of web sites so that, for example, database such as that noted in Dworkin, can be populated with information in a cost effective, efficient manner, without requiring creation of crawlers for each different web site. In contrast to Dworkin which is a front end system and method for facilitating the providing of product information from a database to a consumer, the present invention is a back end system and method for extracting desired data from a plurality of web sites for use in populating such databases.

To perform the functions described, the present method as defined in independent claim 1 specifically recites, *inter alia*:

> (ii) developing an extraction pattern based on <u>output from the respective web site</u> using a graphical user interface tool, the extraction pattern identifying at least a portion of the <u>output of a web site</u> and extracting information from the respective web site W; and

> (iii) associating the developed extraction pattern with the respective description of data of interest for the web site W;

> (B) receiving a value that can be used as an extraction parameter for the developed extraction patterns; and

(C) obtaining said data of interest by querying web sites in the plurality of web sites using the value and the extraction patterns associated with the respective descriptions of data of interest. (Emphasis added.)

Similarly, independent claim 18 specifically recites, *inter alia*:

(ii) means for developing an extraction pattern based on output from the web site using a graphical user interface tool, the extraction pattern extracting data from the output of the web site; and

(iii) means for associating the developed extraction pattern with the respective description of data of interest for the web site W;

(B) means for receiving a value that can be used as an extraction parameter in the developed extraction patterns.

(C) means for obtaining said data of interest by querying web sites in the plurality of web sites using the value and the developed extraction patterns associated with the respective descriptions of data of interest.

Moreover, independent claim 32 specifically recites, *inter alia*:

(A) developing a description of data of interest for each web site in said plurality of web sites based on output from the plurality of web sites using a graphical user interface tool that includes a web browser, each respective description of data of interest including an extraction pattern identifying at least a portion of the output of a web site and extracting user specified information from the corresponding web site;

(B) receiving a value that can be used as an extraction parameter for the developed extraction patterns; and

(C) obtaining said data of interest by querying web sites in the plurality of web sites using the value and the extraction patterns in the respective descriptions of data of interest.

Correspondingly, each of the independent claims 1, 18, and 32 recite developing an extraction pattern based on the <u>output of the web site</u> using a graphical user interface tool. In addition, these claims further recite obtaining data of interest by querying web sites using a received value corresponding to an extraction parameter, and the developed extraction patterns.

It should be understood that the recited "extraction pattern" are portions or components of the "description of data of interest" which are used to extract the data of interest from a web site. In the recited claims, the extraction pattern is developed based on output of the web site by using a graphical user interface tool. (See Pg. 3, lines 14-17; Pg. 5, lines 1-5; Pg. 8, lines 19-23; Pg. 15, lines 8-12; Pg. 24, line 23-Pg. 25, line 18; Figs. 4, 16 and 17). Examples of extraction patterns and development thereof are most clearly shown in the implementation of the present invention as shown in Figures 15 to 17, and the corresponding portions of the specification discussing these figures in Page 24, line 23 to Page 25, line 24.

Dworkin does not disclose, teach, or otherwise suggest development of an extraction pattern recited in the rejected claims of the present application. Dworkin does disclose a template that is displayed, a user inputting criteria into the template, and the system searching the database based in the inputted criteria. In the Advisory Action, the Examiner asserts that this provision of the template discloses the recited description of data of interest. (See Advisory Action mailed May 10, 2005, item 13). It may be argued that the inputted criteria disclosed in Dworkin is analogous to the received value recited in the present independent claims 1, 18, and 32. However, Dworkin does not disclose <u>an extraction pattern developed based on output of a web site</u>, the extraction pattern being adapted to extract information from the respective web site as specifically recited in the rejected claims. Dworkin specifically recites that "the term 'template' means a screen display which is analogous to a questionnaire. That is, the template lists certain general features of the product selected, and provides areas in which the user can fill in desired specifications." (See Dworkin, Col. 5, lines 46-50).

It cannot be reasonably argued that the template of Dworkin is equivalent to the extraction pattern recited in the present claim because the template is predefined and provided to the user. More importantly, the template disclosed in Dworkin <u>is not developed based on the output</u> of the respective web site. In this regard, Dworkin is silent as to how the template is initially derived. Dworkin discloses that the <u>database is queried</u> using the template and the inputted criteria, but presumes that the template will work with the database, and function to obtain the desired information from an

existing database that is already populated with product information. (See Dworkin, Col. 5, line 55-Col. 6, line 15). Correspondingly, without specific teaching to the contrary, it would be evident to one of ordinary skill in the art that Dworkin merely discloses conventional searching techniques using declarative queries of a structured query language to search an existing populated database.

In contrast, in the present invention specifically recited in independent claims 1, 18, and 32, the extraction patterns are developed based on the output of the web site itself. Thus, the extraction patterns are not predefined, at least not until they are developed using the output of the web site. In addition, the rejected claims further recite that <u>a plurality of web sites</u> are queried using the value and the developed extraction patterns to extract the data of interest. Conventional structured query language cannot be readily used to extract desired information from web sites, for example, because most web sites include a plurality of web pages that need to be navigated to obtain the desired information. Moreover, claims 1 and 18 also specifically recite that when the data of interest includes data from two or more web sites, the data from the web sites are provided. Such feature is not disclosed or suggested by Dworkin because the system of Dworkin works exclusively within the database provided.

The Examiner asserts that the present invention sets forth an HTML web product and service search engine tool using standard database software tools and programming software tools. (See Final Office Action of October 5, 2004, page 2, item 2). This assertion is incorrect in that, as discussed above, the present invention is uniquely provided for allowing extraction of data of interest from a web site <u>by facilitating development of extraction patterns using the output of the web site itself.</u> As described, the present invention provides a graphical user interface tool to facilitate development of the extraction patterns. Once an extraction pattern has been developed for a web site based on the output of the web site itself, new values indicative of the desired data of interest are used in conjunction with the developed extraction pattern, to extract the desired information from the web site.

In view of the above, the Applicants respectfully contend that the Examiner's reliance on Dworkin and summary assertions as to obviousness based on databases

and programming languages is improper, and does not establish a *prima facie* case of obviousness. These assertions of obviousness are made by the Examiner without properly establishing any basis for modifying prior art systems and methods such as that disclosed in Dworkin to provide the features and functionality recited in the rejected claims. In this regard, Examiner appears to be engaging in improper hindsight reconstruction based on the present invention to obtain the required motivation for combining the references or teachings to assert that the present invention is "obvious", without properly citing any teachings, or establishing proper motivation.

The impropriety of the Examiner's summary rejections is evidenced by the fact that the Examiner has taken twelve "official notices" in rejecting the pending claims as being obvious, when use of such notices should be rare and judiciously applied. (See MPEP 2144.03). For example, the Examiner asserts that Dworkin teaches providing a tool for creating a program to extract data using at least one extraction parameter. While the Examiner further admits that Dworkin does not teach the web site, he takes official notice that link construction is well known and that it would be obvious to implement this feature "for the advantage of increased revenue by greater exposure to on-line customers and products." This statement reveals that the Examiner is not fully appreciating the present invention which is directed to extraction of information from web sites, a function that cannot be performed by the system of Dworkin without substantial modifications thereto. Again, as discussed above, Dworkin merely disclose a database search tool for searching a pre-existing database, whereas the present claims recite developing extraction patterns using the output of the web site itself. Thus, the recited present invention is not suggested by Dworkin, by web sites generally, or by link construction in HTML. The Examiner merely taking notice of existence of web sites does not address the deficiencies of the rejection in that even if Dworkin is modified with the teaching officially noted by the Examiner, the modified Dworkin reference still fails to result in the present invention.

Therefore, the Applicants respectfully contend that the Examiner's rejection of independent claims 1, 18, and 32, as well as dependent claims 2, 20, 25, 26, 29, 30, and 34 of Group I is improper.

B.   GROUP II

Claim 4 which is the sole claim in Group II, is dependent on independent claim 1 discussed *supra*, but further recites, *inter alia*:

> applying the extraction pattern to the output of the web site that is displayed in a source view in the web browser thereby identifying the at least a portion of the output for the web site; and
>
> displaying a rendered version of the at least a portion of the output of the web site.

Therefore, claim 4 discloses the method recited in claim 1, further including displaying the web site in a source view, and displaying a rendered version of at least a portion of the output of the web site. (See Pg. 15, lines 10-13; Pg. 16, lines 7-8). Displaying the web site in a source view facilitates application of the extraction pattern. Because the cited Dworkin reference fails to teach or otherwise suggest a system or method including an extraction pattern, or extraction of data from web sites, the Applicants respectfully contend that Dworkin also fails to render obvious, the invention recited in Group II, and that the Examiner's rejection of dependent claim 4 is improper.

C.   GROUP III

Claims of Group III include claims 5, 6, 9, 10, 16, and 19 are directed to providing a listing of extraction patterns or commands that can be selected by the user. Providing of such listing would facilitate and expedite creating respective descriptions of data so that extraction of data from a plurality of web sites can be attained in still a more cost effective manner. As discussed *supra*, Dworkin fails to disclose, teach, or otherwise suggest extraction patterns, much less providing a listing of such extraction patterns to further facilitate extraction of data of interest from a plurality of websites. Thus, the Applicants respectfully contend that the Examiner's rejection of dependent claims 5, 6, 9, 10, 16, and 19 of Group III is improper.

D.   GROUP IV

Group IV with claim 7 is directed to a method of extracting data of interest from a plurality of web sites and further requires identifying, and submitting a form in the output of the web site. The recited limitation "form" refers to a search field provided directly on the output of a web site, an example being identified with numeral 900 in Figure 9. (See Pg 11, lines 4-6; Pg. 23, lines 3-4). Recited forms allow particular products to be identified in the web site, and is a feature provided on the web site itself. As discussed *supra*, Dworkin fails to disclose, teach, or otherwise suggest, a system or method of extracting data of interest from a plurality of web sites. Correspondingly, Dworkin also fails to teach, or otherwise suggest identifying a form in a web site, or submitting the form, as specifically recited in claim 7. Therefore, the Applicants contend that Examiner's rejection of claim 7 is improper.

E.   GROUP V

Group V with claim 11 is directed to a method of extracting data of interest from a plurality of web sites and further requires:

> an extraction command for extracting data from a first web site
> and a second web site, the first web site including a reference to the second
> web site.

Thus, claim 11 specifically recites linking of a first web site with a second web site, and extraction of data from both the first web site and the second website. As discussed *supra*, Dworkin fails to disclose, teach, or otherwise suggest, a system or method of extracting data of interest from a plurality of web sites, much less web sites that are linked together. Therefore, the Applicants contend that Examiner's rejection of claim 11 is improper.

F.   GROUP VI

Claims 12-15, 17, 21, and 31 of Group VI recite a test condition to test instructions associated with an extraction pattern using a user input. (See Pg. 19, lines 3-6; Pg. 24, lines 15-17; Figs. 4 and 13). The cited Dworkin reference fails to teach,

or otherwise suggest a system or method including test conditions for testing instructions associated with an extraction pattern using a user input. There would be no motivation for providing of such testing and such testing would be unnecessary in the system and method disclosed in Dworkin because the data structure of the database of Dworkin would be known, and the templates are provided to search merely within the provided database. Correspondingly, because Dworkin does not disclose or suggest development of an extraction pattern or templates, the desirability or motivation for providing testing is eliminated. Therefore, the Applicants respectfully contend that the Examiner's rejection of the claims of Group VI is improper.

G. GROUP VII

The rejected claims 23, 24, 27, 28, and 33 of Group VII are directed to the types of expressions provided in an extraction pattern. In this regard, exemplary claim 23 recites:

> wherein said extraction pattern comprises a pre-condition regular expression, a portion of data of interest regular expression, and a post-condition regular expression and wherein said developing comprises refining at least one of said pre-condition regular expression, said portion of data of interest regular expression, and said post-condition regular expression.
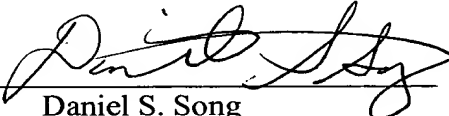
As discussed *supra*, Dworkin fails to disclose, teach, or otherwise suggest extraction patterns as claimed. Correspondingly, Dworkin also fails to teach, or otherwise suggest the types of expressions provided in an extraction pattern. Moreover, these claims further recite refining of at least one of these expressions. There would be no motivation for providing of such refining in Dworkin, and such testing would be unnecessary in the system and method of Dworkin as well because the data structure of the database of Dworkin would be known, and the templates used therein are provided to search merely within the provided database. Therefore, the Applicants respectfully contend that the Examiner's rejection of the claims of Group VII is also improper.

X.   CONCLUSION

Thus, at least for the foregoing reasons, the Applicants contend that the applied Dworkin reference does not render the claimed invention obvious and unpatentable. The reversal of the Examiner's rejection under 35 U.S.C. §103 with respect to all of the pending claims 1, 2, 4-7, 9-21, and 23-34 of the present application is respectfully requested.

Respectfully submitted,
NIXON PEABODY, LLP

Date: **June 13, 2005**

Daniel S. Song
Registration No. 43,143

Customer No.: 22204

NIXON PEABODY LLP
401 9th Street, N.W., Suite 900
Washington, DC 20004
(202) 585-5000
(202) 585-8080 (Fax)

APPENDIX

1. A method of extracting data of interest from a plurality of web sites, the method comprising:

(A) for each respective web site W in said plurality of web sites,

(i) creating a respective description of data of interest that identifies the web site W;

(ii) developing an extraction pattern based on output from the respective web site using a graphical user interface tool, the extraction pattern identifying at least a portion of the output of a web site and extracting information from the respective web site W; and

(iii) associating the developed extraction pattern with the respective description of data of interest for the web site W;

(B) receiving a value that can be used as an extraction parameter for the developed extraction patterns; and

(C) obtaining said data of interest by querying web sites in the plurality of web sites using the value and the extraction patterns associated with the respective descriptions of data of interest, wherein

when the data of interest includes data of interest from at least two web sites of the plurality of web sites, the data of interest from the at least two web sites is provided.

2. The method of claim 1, wherein the graphical user interface tool includes a web browser.

3. (Canceled)

4. The method of claim 2, further comprising:

applying the extraction pattern to the output of the web site that is displayed in a source view in the web browser thereby identifying the at least a portion of the output for the web site; and

displaying a rendered version of the at least a portion of the output of the web site.

5. The method of claim 2, wherein the graphical user interface tool further includes a plurality of predefined extraction patterns.

6. The method of claim 5, wherein the plurality of predefined extraction patterns includes at least one of an extraction pattern for matching a hyperlink, an extraction pattern for matching a form, and an extraction pattern for matching a price.

7. The method of claim 2, wherein the graphical user interface tool further comprises:

identifying a form in the output of the web site;

creating a step in the description of data of interest corresponding to the web site, the step to submit the form without retrieving the web site;

generating a plurality of parameters associated with the step, each parameter in said plurality of parameters corresponding to an input in the form; and

associating a parameter in the plurality of parameters with the extraction parameter.

8. (Canceled)

9. The method of claim 1, wherein the developing of an extraction pattern includes receiving a selection of an extraction command from a predetermined list of extraction commands.

10. The method of claim 9, wherein the predetermined list of extraction commands includes an extraction command for retrieving multiple matches of an extraction pattern from a web site.

11.   The method of claim 9, wherein the predetermined list of extraction commands includes an extraction command for extracting data from a first web site and a second web site, the first web site including a reference to the second web site.

12.   The method of claim 9, wherein at least one step in the plurality of steps includes a test condition comprising a logical test for at least one corresponding argument and a first step in the plurality of steps, and wherein the respective description of data of interest continues executing at the first step when the logical test is satisfied.

13.   The method of claim 12, wherein the at least one corresponding argument includes an extraction pattern.

14.   The method of claim 12, wherein the test condition further comprises a result code that returns an error when the output of the respective web site has changed.

15.   The method of claim 12, wherein the test condition further comprises a result code that returns an error when the output of the respective web site has no information about the product.

16.   The method of claim 9, wherein the predetermined list of extraction commands includes an extraction command for segmenting the output of the respective web site into a plurality of units, each of the plurality of units matching the extraction pattern.

17.   The method of claim 16, wherein developing an extraction pattern includes using an extraction command to segment the output of the respective web site into a plurality of units, and using a test condition that comprises a logical test and at least one argument, and wherein for each of the plurality of units, the logical test is computed with the at least one argument, and the unit is removed from the plurality of units if the logical test is not satisfied with the at least one argument.

18. An apparatus for extracting information of interest from a plurality of web sites, the apparatus comprising:

(A) for each respective web site W in the plurality of web sites,

(i) means for creating a respective description of data of interest that identifies the web site W;

(ii) means for developing an extraction pattern based on output from the web site using a graphical user interface tool, the extraction pattern extracting data from the output of the web site; and

(iii) means for associating the developed extraction pattern with the respective description of data of interest for the web site W;

(B) means for receiving a value that can be used as an extraction parameter in the developed extraction patterns; and

(C) means for obtaining said data of interest by querying web sites in the plurality of web sites using the value and the developed extraction patterns associated with the respective descriptions of data of interest,

wherein, when the data of interest includes data from at least two web sites of the plurality of web sites, the means for providing product information provides the data of interest from the at least two web sites.

19. The apparatus of claim 18, wherein the means for developing an extraction pattern includes means for selecting an instruction from a predetermined list of instructions.

20. The apparatus of claim 18, wherein the graphical user interface tool that comprises a web browser.

21. A computer data signal embodied in a carrier wave comprising:

(A) a software module for creating a description of data of interest, the software module including;

(i) a set of operations for interactively developing an extraction pattern based on output of a target web site using a graphical user interface tool, the developed extraction pattern for obtaining data of interest from the target web site;

(ii) a set of operations for receiving a selection of an instruction from a predefined set of instructions for inclusion in the description of data of interest;

(iii) a set of operations for associating the extraction pattern with the instruction;

(iv) a set of operations for testing the instruction using the extraction pattern and the contents of a buffer, wherein the buffer includes a portion of the output of the web site associated with the description of data of interest; and

(B) a software module for using said description of data of interest to obtain data of interest from the target web site when a value that can be used as an extraction parameter for the developed extraction pattern is provided.

22. (Canceled)

23. The method of claim 1 wherein said extraction pattern comprises a pre-condition regular expression, a portion of data of interest regular expression, and a post-condition regular expression and wherein said developing comprises refining at least one of said pre-condition regular expression, said portion of data of interest regular expression, and said post-condition regular expression.

24. The method of claim 23 wherein said portion of data of interest regular expression includes a variable that is replaced with said value for said extraction parameter during said providing.

25. The method of claim 1 wherein the data of interest is provided incrementally as it is obtained from the plurality of web sites.

26. The method of claim 1 wherein, the data of interest is obtained from the plurality of web sites and then presented simultaneously.

W656626.1

27.   The apparatus of claim 18 wherein said extraction pattern comprises a pre-condition regular expression, a portion of data of interest regular expression, and a post-condition regular expression and wherein said means for developing comprise refining at least one of said pre-condition regular expression, said portion of data of interest regular expression, and said post-condition regular expression.

28.   The computer data signal of claim 21 wherein said extraction pattern comprises a pre-condition regular expression, a portion of data of interest regular expression, and a post-condition regular expression and wherein said operations for developing comprise refining at least one of said pre-condition regular expression, said portion of data of interest regular expression, and said post-condition regular expression.

29.   The method of claim 1 wherein the data of interest is information associated with a product or information associated with a service.

30.   The apparatus of claim 18 wherein the data of interest is information associated with a product or information associated with a service.

31.   The computer data signal of claim 21 wherein said data of interest is a product, information, or a service.

32.   A computer implemented method of obtaining data of interest from a plurality of web sites comprising:

(A) developing a description of data of interest for each web site in said plurality of web sites based on output from the plurality of web sites using a graphical user interface tool that includes a web browser, each respective description of data of interest specifying an address for a corresponding web site in the plurality of web sites and each respective description of data of interest including an extraction pattern identifying at least a portion of the output of a web site and extracting user specified information from the corresponding web site;

(B) receiving a value that can be used as an extraction parameter for the developed extraction patterns; and

(C) obtaining said data of interest by querying web sites in the plurality of web sites using the value and the extraction patterns in the respective descriptions of data of interest.

33. The computer implemented method of claim 32 wherein each said extraction pattern comprises a pre-condition regular expression, a portion of data of interest regular expression, and a post-condition regular expression and wherein said developing comprises refining at least one of said pre-condition regular expression, said portion of data of interest regular expression, and said post-condition regular expression.

34. The computer implemented method of claim 32 wherein said data of interest is a product, information, or a service.